

早稲田大学 人間科学学術院 人間科学会 諸費用補助成果報告書（Web 公開用）

申請者（ふりがな）	大浦 杏奈 （ おおうら あんな ）
所属・資格（※学生は課程・学年を記載。卒業生・修了生は卒業・修了年月も記載）	修士課程 2 年
発表年月 または事業開催年月	2024 年 3 月
発表学会・大会 または事業名・開催場所	日本音響学会第 151 回(2024 年春季)研究発表会
発表者（※学会発表の場合のみ記載、共同発表者の氏名も記載すること）	大浦杏奈、菊池英明、小林哲則
発表題目（※学会発表の場合のみ記載）	音声信号からの real-time MRI 調音運動動画の推定
発表の概要と成果（抄録を公開している URL がある場合、「概要・成果」を記載した上で、URL を末尾に記してください。また、抄録 PDF は別途ご提出ください。なお、抄録 PDF は Web 上には公開されません。）	
【概要】 音声信号から調音運動を推定する技術（Acoustic-to-articulatory inversion; AAI）の開発は、古くから取り組まれている。1970 年代は、声道形状を表現する調音モデルを事前に準備し、その調音モデルのパラメーターを推定する AAI が一般的であった。一方で、近年は機械学習を用いて、音声信号から real-time MRI (rtMRI) 調音運動動画を直接推定する AAI が開発されている。 rtMRI 動画推定に関する先行研究では、英語発話時の rtMRI 動画を対象としていた。そのため、日本語発話における調音運動も推定可能かを検討するためには、既存モデルを日本語音声 rtMRI 動画に適応する必要がある。また、先行研究では、入力データの前処理を行っていない点を課題として挙げている。加えて、先行研究では LSTM を用いたモデルによる検討が実施されていたが、LSTM を双方向に行う Bidirectional LSTM (BLSTM) を用いることで、さらに滑らかな推定 rtMRI 動画の出力が可能になると期待される。 そこで本研究では、先行研究モデルの日本語音声 rtMRI 動画への適応に取り組み、英語音声 rtMRI 動画の結果との比較を行った。加えて、前処理として正規化処理の提案と、既存の LSTM モデルの一部を変更した BLSTM モデルにおける推定精度の検討も実施した。 【成果】 日本語発話を対象に推定 rtMRI 動画を出力した結果、以下 3 つの考察が得られた。 - 英語発話を対象とした先行研究と同程度の精度で、日本語発話の推定が可能である。したがって、rtMRI 動画における言語によらずに調音運動を推定することが可能である。 - 前処理として正規化処理を施すことで、顔の位置がより安定した推定が可能である。 - BLSTM では LSTM より滑らかな rtMRI 動画の推定が可能になるものの、正規化処理による画像の歪みの影響を受けやすい可能性がある。 以上の成果を踏まえて、今後はより高精度な推定が可能な AAI の開発のために、Vision Transformer などの最新の手法を用いたモデルの構築に取り組む予定である。 （抄録 PDF を別添。）	

※無断転載禁止